Single nucleus transcriptomics data integration recapitulates the major cell types in human liver

Hepatology Research

Hepatology Research 2020 doi: 10.1111/hepr.13585

Short Communication

Single nucleus transcriptomics data integration recapitulates the major cell types in human liver

Klev Diamanti,[1] Juan Salvador Inda Díaz,[2] Amanda Raine,[3] Gang Pan,[1] Claes Wadelius[1] and Marco Cavalli[1]

Science for [1]Life Laboratory, Department of Immunology, Genetics and Pathology, and [3]Life Laboratory, Department of Medical Sciences, Molecular Medicine, Uppsala University, Uppsala, [2]Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden

Aim: The aim of this study was to explore the benefits of data integration from different platforms for single nucleus transcriptomics profiling to characterize cell populations in human liver.

Methods: We generated single-nucleus RNA sequencing data from Chromium 10X Genomics and Drop-seq for a human liver sample. We utilized state of the art bioinformatics tools to undertake a rigorous quality control and to integrate the data into a common space summarizing the gene expression variation from the respective platforms, while accounting for known and unknown confounding factors.

Results: Analysis of single nuclei transcriptomes from both 10X and Drop-seq allowed identification of the major liver cell types, while the integrated set obtained enough statistical power to separate a small population of inactive hepatic stellate cells that was not characterized in either of the platforms.

Conclusions: Integration of droplet-based single nucleus transcriptomics data enabled identification of a small cluster of inactive hepatic stellate cells that highlights the potential of our approach. We suggest single-nucleus RNA sequencing integrative approaches could be utilized to design larger and cost-effective studies.

Key words: 10X, data integration, Drop-seq, liver, snRNA-seq

INTRODUCTION

THE HUMAN LIVER represents a heterogeneous but well-defined tissue composed of parenchymal cells (PCs) and non-PCs. Hepatocytes (HCs) constitute the PCs and the largest part of the organ, and are involved in diverse metabolic processes such as drug metabolism, bile acid synthesis, and lipid metabolism. Non-PCs regulate HCs through signaling factors they release.[1] Non-PCs include, among others, liver sinusoidal endothelial cells (LSECs), macrophages known as Kupffer cells (KCs) that reside in the sinusoid capillary lumen, and hepatic stellate cells (HSCs) that act as storage depots for fat and vitamin A when in quiescent form, and undergo a morphological transformation into myofibroblastic cells upon activation following liver injury.[2]

The study of single-cell transcriptomes (scRNA-seq) provides detailed information on the heterogeneity of tissues and organs, unveils novel or rare cell populations or subpopulations, allows tracking developmental trajectories and cell lineages, and potentially leads to new biological and clinical insights.[3–5] After the

Correspondence: Dr. Marco Cavalli, Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Husargatan 3C, Box 815, 751 08 Uppsala, Sweden. Email: marco.cavalli@igp.uu.se
Conflict of interest: The authors have no conflict of interest.
Financial support: The study was supported by grants from SciLifeLab, AstraZeneca, the Swedish Diabetes Foundation (DIA 2017-269), and EXODIAB. Received 17 August 2020; revision 2 October 2020; accepted 20 October 2020.
Received 17 August 2020; revision 2 October 2020; accepted 20 October 2020.

Correspondence: Dr. Marco Cavalli, Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Husargatan 3C, Box 815, 751 08 Uppsala, Sweden. Email: marco.cavalli@igp.uu.se

pioneering report on the exploration of single cell transcriptomes more than a decade ago,[6] multiple methodologies have been developed.[3,5,7] Selection of the suitable scRNA-seq method should be primarily based on evaluation of the nature of the samples and the costs. Other key parameters to take into account include the correspondence between the quantification of sequencing reads and mRNA concentration (accuracy), and the probability of capturing a specific mRNA transcript (sensitivity).[3] The biological question (s) driving the design of the experiment should aim at retaining the balance between the qualitative and quantitative aspects of the methodology.

A major bottleneck in carrying out scRNA-seq in tissue samples is posed by the limited availability of fresh tissue and, most often, by the technical challenges in isolating rare cell types.[8] The majority of biobanks store flash-frozen tissue samples that subsequently do not facilitate isolation of viable single cells for scRNA-seq. However, several studies have shown that the transcriptomics profile of intact single nuclei closely resembles the one of the original cells.[8–10] Droplet microfluidics methods designed for scRNA-seq profiling have been adapted to carry out single-nucleus RNA sequencing (snRNA-seq).[9,10] Popular droplet-based approaches suitable for snRNA-seq include Chromium 10X Genomics[11] and Drop-seq.[12] Although the overall strategy of both methodologies involves single nuclei encapsulation in droplets, cell barcoding to track cells, and unique molecular identifiers (UMIs) to count unique mRNAs, they differ in the cDNA amplification and in-droplet reactions.[7,13] Overall, 10X offers better sensitivity and less technical noise, albeit at greater monetary cost.[13] Several technical differences between the protocols could potentially account for the increased resolution of 10X over Drop-seq. For instance, in 10X the reverse transcription is carried out inside the small volume of the droplets, whereas in Drop-seq the beads, with captured mRNA molecules on the surface, are released from the emulsion before the reverse transcription is carried out in a bulk reaction.

Transcriptomics analysis of human liver tissue at single cell level has recently painted a detailed view of the organ design.[14–17] Integrating and exploring single nuclei transcriptomes from different technologies could help uncover cell subpopulations previously indecipherable in individual experiments due to resolution or power limitations. Another advantage of data integration is the increased number of features in the integrated space that enables higher accuracy for cluster and marker identification. Here, we integrated snRNA-seq from a human liver tissue sample using 10X and Drop-seq, and we explored cell populations in the integrated space. The integration of single nuclei transcriptomics profiles with differing resolution recapitulated the major cell types of the human liver, and increased the statistical power that allowed identification of cell populations not identified in the respective experiments. Finally, we discussed the benefits and the caveats of the integrative approach.

METHODS

Ethics statement

THE USE OF the liver tissue sample used in this study was approved by the Uppsala regional ethics committee (Dnr: 2014/433).

Nuclei preparation

The human liver tissue sample was provided by Professor Per Artursson at Uppsala University. The sample was obtained from a partial hepatectomy of a male patient with colon cancer and hepatic metastasis. The patient had provided written consent for the use of the biological sample for scientific research. Part of the resection, characterized as tumor-free by a pathologist, was flash-frozen and subsequently used in this study.

Nuclei suspension was prepared using the gentleMACS dissociator (Miltenyi Biotec, Bergisch Gladbach, Germany) as described earlier by Cavalli *et al.*[17] The liver nuclei were resuspended in 2 mL phosphate-buffered saline with 0.01% bovine serum albumin for the Drop-seq preparation and RiboLock RNase Inhibitor (Thermo Fisher Scientific, Waltham, MA, United States) and filtered using 30 μm MACS SmartStrainers. The nuclei suspension was evaluated for purity and concentration using the Countess II Automated Cell Counter (Thermo Fisher Scientific, Waltham, MA, United States) by staining the nuclei with Trypan blue and was adjusted to a final concentration of ~300 nuclei/μL for the Drop-seq using the Nadia instrument (Dolomite Bio, Royston, Hertfordshire, UK).

cDNA library preparation and sequencing

Single nuclei transcriptomics data for Chromium 10X was obtained from Cavalli et al.[17] Drop-seq was carried out using the Nadia instrument (Dolomite Bio) for nuclei droplet encapsulation as previously described,[12] with the exception that emulsion breakage was undertaken by filtration through a 5-μM uberstrainer (pluriSelect, Leipzig, Germany). cDNA amplification was carried out by polymerase chain reaction (4 + 13 cycles) using aliquots of 10 000 beads. Three cDNA amplification reactions,

corresponding to 30 000 beads, were pooled and purified twice by 0.6X AMPure beads (Beckman Coulter Life Sciences, Indianapolis, IN, United States) and 600 pg of amplified DNA was used as input for Nextera library (Illumina, San Diego, CA, United States) preparation. Libraries were sequenced using a custom sequencing oligo[12] on a Illumina NovaSeq SP flow cell as follows: read 1, 25 bp; read 2, 91 bp; and index read, 8 bp.

Data preprocessing, nuclei clustering, and differential expression

The raw base call files from the sequencer for the 10X experiment were demultiplexed using the *mkfastq* function of the *cellranger* tool version 2.2 provided by 10X Genomics. Raw base call files from the Drop-seq experiment were converted to fastq using the tool *bcl2fastq* version 2.20 by Illumina. Fastq files from 10X and Drop-seq were imported in the *dropSeqPipe* pipeline (https://github.com/Hoohm/dropSeqPipe) for alignment, quantification, and calculation of the UMI matrices (Table S1).

Unfiltered matrices containing UMIs for both experiments were used for the downstream analysis (Fig. S1). We applied debris identification using expectation maximization (DIEM) to identify empty/contaminated droplets and assess the quality of droplets that otherwise "appear healthy".[18] Droplets containing less than 200 UMIs or five genes were assigned to the predefined debris set, while the rest were included in the test set (Fig. S2). DIEM measures the distance of clusters of droplets from the test set to those of the debris set. The minimum distance that any of the $k = 30$ clusters from the test set is required to have from the debris set was 0.2 (Fig. S3). This filtering process resulted in 1386 nuclei for Drop-seq and 2475 nuclei for 10X.

We used the *SCTransform* function from *Seurat* as a first normalization approach on both datasets, controlling for mitochondrial percentage.[19,20] We selected 3592 genes with residual variance above 1.3. Next, we used the raw mRNA counts and ZINB-WaVE, a zero-inflated negative binomial model developed for scRNA-seq to integrate single nucleus datasets originating from different technologies.[21] A matrix of ($K = 20$) surrogate variables that represented unwanted variation in the dataset was calculated using the 3592 genes from the previous step, the confounding factors included were the snRNA-seq methodology (Drop-seq and 10X) and the log-transformed mRNA counts per cell, and a general regularization parameter (*epsilon = 1000*). The same matrix was used to construct a shared nearest neighbors graph and cluster the cells using the Louvain algorithm (functions *FindNeighbors* and *FindCluster* from the R package *Seurat*). The average silhouette width for different resolution levels in the clustering algorithm was calculated (Fig. S4a) and six clusters (HC, KC, LSEC, quiescent HSC [qHSC], active HSC [aHSC], and inactivate HSC [iHSC]) were obtained for resolution *0.3* in the integrated set (Figs S4,S5). Average silhouette width is used to assess the quality of the clustering, and it ranges between −1 (poor clustering) and 1 (good clustering). Clusters did not show notable biases due to cell cycle, number of genes, or number of UMIs (Fig. S5).

We repeated the dimensionality reduction and clustering on the original Drop-seq and 10X, with the difference of using the top 1000 variable genes from each and the log values of mRNA counts per cell as covariates on each dataset. For 10X, five clusters (HC, KC, LSEC, qHSC, and aHSC) were defined at resolution *0.5*, while four clusters (HC, KC, LSEC, and aHSC) were identified in Drop-seq for resolution *0.4* (Figs S4,S5). Similar to the integrated set, these datasets did not show any noteworthy influence of cell cycle, number of genes, or number of UMIs (Fig. S5). The adjusted Rand index between the clusters for Drop-seq and 10X was 0.84 and 0.82, respectively. This indicates that the integration does not distort the general structure of the datasets. Differential expression analysis among clusters was calculated with a Wilcoxon test on the raw counts and visualized on a heatmap for the SCTransform data for the top 10 differentially expressed genes and cell type markers (Fig. S6).

Data availability

All relevant data are presented within the article and its supporting information files. Raw data is available at the European Nucleotide Archive under the EBI BioStudies accession number S-BSST324. Specifically, the accession numbers for the experiments are ERX3897748 and ERX4298533. These data will be also available from the corresponding author upon request.

RESULTS AND DISCUSSION

In the present study we integrated transcriptomics data from human liver obtained from two droplet-based techniques for high-throughput snRNA-seq, 10X and Drop-seq.[17] Raw data underwent extensive quality control, resulting in 2475 and 1386 nuclei for 10X and Drop-seq, respectively (Table S1). Finally, the set of 3861 nuclei from the two methodologies was integrated using ZINB-WaVE and further downstream analysis was carried out (see "Methods").[21]

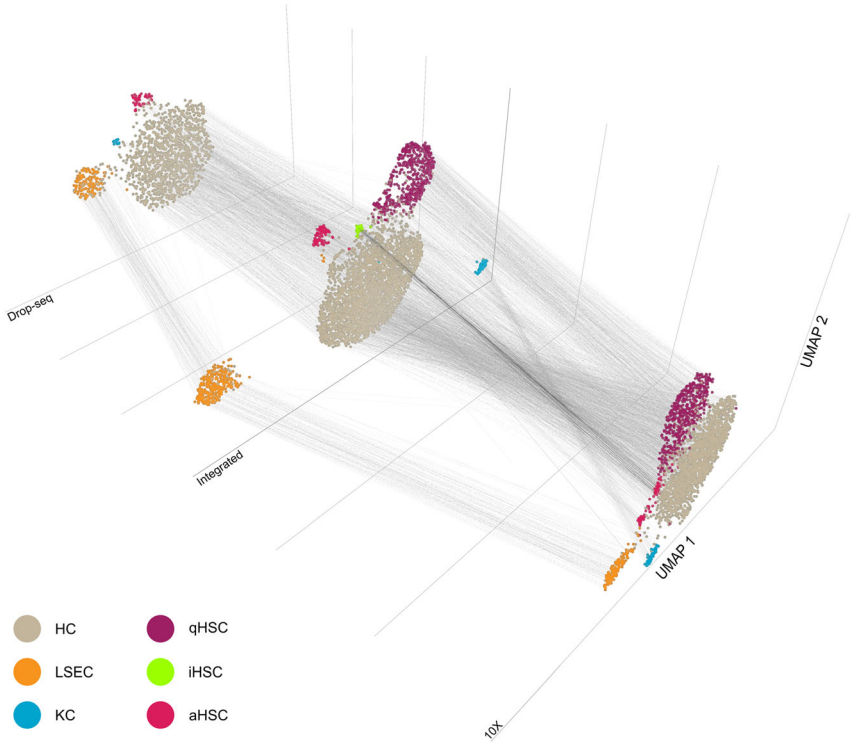

Figure 1 UMAP plots for Drop-seq, Chromium 10X Genomics, and the integrated set of single-nucleus RNA sequencing data for a human liver sample (Fig. S5). Connections from Drop-seq and 10X towards the integrated set represent contributing nuclei to the clusters of the integration. All connections to inactivate hepatic stellate cells (iHSC) are in black; the remaining connections are in gray. An interactive version of the figure is also available (Data S1). aHSC, active HSC; HC, hepatocytes; KC, Kupffer cells; LSEC, liver sinusoidal endothelial cells; qHSC, quiescent HSC.

Fine-tuning the clustering parameters in combination with correcting for various confounding factors revealed six distinct cell populations that were subsequently annotated based on the expression of cell-type specific markers curated from other relevant liver studies (see "Methods"; Figs S1–S4).[14–16,22,23] Specifically in the integrated set we identified 2761 HCs, 305 LSECs, 82 KCs, and three clusters of HSCs representing different activation states, namely 584 qHSCs, 94 aHSCs, and 35 iHSCs (Figs 1,S4–S6).

HCs were the most abundant cell type in individual experiments, as well as in the integrated set. They were annotated by the expression of well-known marker genes including *G6PC*, *PCK1*, *PLG*, *FGL1*, *C3*, and *F5*, and several genes encoding for proteins in the cytochrome P450 family or in the complement cascade (Figs 1,S7, Table S2).

Liver sinusoidal endothelial cells were also identified from both 10X and Drop-seq independently and they formed a distinct cluster in the integrated space. The LSEC marker genes included *FLT1*, *STAB1*, *OIT3*, *AKAP12*, *PTPRB*, *NOSTRIN*, and *PLEKHG1* (Figs 1,S7, Table S2).

Similarly, KCs were defined using both single nucleus transcriptomics approaches and were annotated by the expression of immunity and macrophage/monocyte-specific genes such as *CD163*, *CD74*, *IL18*, *MYO1F*, *FGD2*, *SAT1*, *MARCO*, and *MSR1* (Figs 1,S7, Table S2).

The identification of several clusters of HSCs highlighted the increased resolution that can be achieved by integrating data from various sources. Both 10X and Drop-seq identified a cluster of aHSCs characterized by signature expression of genes such as *PDGFRB*, *ADAMTS2*, *LAMA2*, *LAMB1*, and *DCN* and several genes encoding for collagen, metalloproteinases and other extracellular matrix proteins (Figs 1,S7, Table S2). Quiescent HSCs, defined by genes including *UCHL1*, *NTM*, *NRXN1*, and *PCDH7*, were identified solely by 10X, whereas the lower resolution of Drop-seq prevented the detection of qHSCs, despite a subset of nuclei expressing some of the cluster markers (Fig. 1).

Recovery from liver injury is accompanied by reduction of the fibrosis that is characterized by the disappearance of aHSCs, which can be eliminated through apoptosis or

enter a senescence state that prompts an immune-mediated removal. Recent experiments in mice have indicated that aHSCs can revert into an inactivated, quiescent-like phenotype once the injury source is removed.[24] Inactivated HSCs display a distinct gene expression profile marked by downregulation of fibrogenic genes (*COL1A1*, *COL1A2*, *ACTA2*, *TGFBR1*, and *TIMP1*) and upregulation of some quiescence-associated genes (*PPARG* and *BAMBI*), but not adipogenic genes such as *ADFP*, *ADIPOR1*, or *GFAP*.[25] Although more similar to qHSCs than to aHSCs, iHSCs do not fully revert to qHSCs, and are primed to be quickly reactivated by reoccurring fibrogenic stimuli.[26,27]

In the integrated space we identified a small cluster of 35 nuclei with a transcriptomics profile bearing gene markers of both qHSCs and aHSCs, likely representing a population of iHSCs. The presence of this cluster was signified by the negative silhouette width for a subset of nuclei from the aHSC in 10X, where the majority of iHSCs originate from (Fig. S4c). As expected, the analysis of the genes defining the cluster revealed lower expression of extracellular matrix proteins with respect to aHSC (Fig. S8a), while the expression of some collagen proteins such as *COL5A3* was retained (Fig. S7). The cluster also showed an increase in expression of quiescence-associated genes with respect to aHSC including *NTM*, *NRXN1*, and *PTN*, but not *ADIPOR1* (Fig. S8b). Finally, the cluster also showed expression of the anti-apoptotic heat-shock protein HSPA1A and HSPA1B that assists iHSCs to avoid apoptotic clearance, in contrast to aHSCs that do not express HSPA1A/B[25] (Table S3).

The iHSCs cluster stems from the aHSCs and originates from nuclei captured with 10X. This cluster was part of aHSC in the 10X dataset (Figs 1,S5), and as a result of the integration with Drop-seq its marker-genes gained power that enabled independent clustering. This is supported by a larger average silhouette width in the case when iHSCs nuclei have their own cluster, compared to when they are grouped together with aHSCs cells (Figs S4,S9).

Single-nucleus RNA sequencing integration allows simultaneous analysis of datasets originating from different samples and/or sequencing technologies. However, it could lead to distorted cell clustering that does not reflect the original grouping of the cells in each individual dataset. Here, we have used integration methodologies developed for scRNA-seq on snRNA-seq. The integrated space showed clusters of nuclei with specific cellular identity that resembled those on each individual dataset. Moreover, snRNA-seq data integration unveiled a cell type masked on the analysis of each dataset separately. To our knowledge, this is the first study to integrate snRNA-seq experiments originating from different technologies (Fig. 1). Analysis of the individual datasets from Drop-seq and 10X showed that the latter contributed largely to the identification of more cell types, whereas the former played an important role to increase the quantity of genes and nuclei. The results from the integrated set highlighted its augmented power to detect smaller subpopulations of cells that ultimately enhance the discovery and interpretability. Guided by this conclusion, we suggest that transcriptomic studies might benefit from combining single cell or nucleus experiments between more costly technologies offering high sensitivity (e.g. Chromium 10X Genomics) and technologies that, due to the limited cost, will allow inclusion of additional samples (e.g. Drop-seq). Integrating data from different sources should enable the discovery of rare/small cell populations in larger cohorts while balancing quality and quantity.

ACKNOWLEDGMENTS

THE STUDY WAS supported by grants from SciLifeLab (CW), AstraZeneca (CW), the Swedish Diabetes Foundation (DIA 2017-269) (CW), and EXODIAB (CW). Single cell sequencing was carried out at the SNP&SEQ Technology Platform at Uppsala University and SciLifeLab. SNP&SEQ is part of the National Genomics Infrastructure (NGI) Sweden supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

REFERENCES

1 Kmiec Z. *Cooperation of liver cells in health and disease: with 18 tables*: Berlin, HeidelbergSpringer, 2001.
2 Arii S, Imamura M. Physiological role of sinusoidal endothelial cells and Kupffer cells and their implication in the pathogenesis of liver injury. *J Hepatobiliary Pancreat Surg* 2000; **7**: 40–8.
3 Ziegenhain C, Vieth B, Parekh S *et al*. Comparative analysis of single-cell RNA sequencing methods. *Mol Cell* 2017; **65**: 631–43.
4 Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018; **36**: 411–20.
5 Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 2018; **50**: 1–4.
6 Tang F, Barbacioru C, Wang Y *et al*. mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009; **6**: 377–82.

7 Mereu E, Lafzi A, Moutinho C *et al.* Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat Biotechnol* 2020; **38**: 747–55.
8 Nguyen QH, Pervolarakis N, Nee K, Kessenbrock K. Experimental considerations for single-cell RNA sequencing approaches. *Front Cell Dev Biol* 2018; **2018**: 6.
9 Habib N, Avraham-Davidi I, Basu A *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* 2017; **2017**(14): 955–8.
10 Gao R, Kim C, Sei E *et al.* Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast cancer. *Nat Commun* 2017; **2017**(8): 228.
11 Zheng GXY, Terry JM, Belgrader P *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017; **8**: 14049.
12 Macosko Evan Z, Basu A, Satija R *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015; **161**: 1202–14.
13 Zhang X, Li T, Liu F *et al.* Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. *Mol Cell* 2019; **73**: 130–42.
14 MacParland SA, Liu JC, Ma X-Z *et al.* Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun* 2018; **9**: 4383.
15 Aizarani N, Saviano A, Sagar *et al.* A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* 2019; **572**: 199–204.
16 Ramachandran P, Dobie R, Wilson-Kanamori JR *et al.* Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* 2019; **575**: 512–18.
17 Cavalli M, Diamanti K, Pan G *et al.* A multi-omics approach to liver diseases: integration of single nuclei transcriptomics with proteomics and hicap bulk data in human liver. *OMICS* 2020; **24**: 180–94.
18 Alvarez M, Rahmani E, Jew B *et al.* Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM. *Sci Rep* 2020; **10**: 11019.
19 Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 2019; **20**: 296.
20 Stuart T, Butler A, Hoffman P *et al.* Comprehensive integration of single-cell data. *Cell* 2019; **177**: 1888–902.
21 Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert J-P. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat Commun* 2018; **9**: 284.
22 Halpern KB, Shenhav R, Massalha H *et al.* Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. *Nat Biotechnol* 2018; **36**: 962–70.
23 Halpern KB, Shenhav R, Matcovitch-Natan O *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* 2017; **542**: 352–6.
24 Troeger JS, Mederacke I, Gwak GY *et al.* Deactivation of hepatic stellate cells during liver fibrosis resolution in mice. *Gastroenterology* 2012; **143**: 1073–83.
25 Kisseleva T, Cong M, Paik Y *et al.* Myofibroblasts revert to an inactive phenotype during regression of liver fibrosis. *Proc Natl Acad Sci USA* 2012; **109**: 9448–53.
26 Tsuchida T, Friedman SL. Mechanisms of hepatic stellate cell activation. *Nat Rev Gastroenterol Hepatol* 2017; **14**: 397–411.
27 Kisseleva T, Brenner DA. Inactivation of myofibroblasts during regression of liver fibrosis. *Cell Cycle* 2013; **12**: 381–2.

SUPPORTING INFORMATION

ADDITIONAL SUPPORTING INFORMATION may be found online in the Supporting Information section at the end of the article.

Data S1 Interactive visualization of Figure 1 in 3D.

Figure S1 Individual knee-plots for Drop-seq and 10X datasets.
Figure S2 Violin plots showing the distribution of number of genes, unique molecular identifiers (UMIs), and mitochondrial genes.
Figure S3 Distances of clusters of nuclei between background and testing sets in DIEM.
Figure S4 Average silhouette value for 10X, Drop-seq, and the integrated set.
Figure S5 UMAP plots for Drop-seq, 10X, and integrated sets.
Figure S6 Expression matrix for the top 10 differentially expressed genes.
Figure S7 UMAP plots for the curated list of cell-type specific markers.
Figure S8 Comparison of marker-gene expression between active hepatic stellate cells (aHSC) and inactive hepatic stellate cells (iHSC).
Figure S9 Silhouette widths for clusters in 10X and the integrated set.

Table S1 Collection of arguments used to run dropSeqPipe for Drop-seq and 10X.
Table S2 Cell-type specific markers used to mark each cluster.
Table S3 Cluster-specific marker genes.